METHOD FOR FABRICATING
MOS DEVICE WITH HALO IMPLANTED REGION

FIELD OF THE INVENTION

The present invention relates to reduced-dimension MOS field effect transistors and to the formation of field effect transistors having narrow gate electrodes and reduced dimension source-drain structures.

5

DESCRIPTION OF THE PRIOR ART

Field-effect transistors, generally referred to as FETs or MOSFETs, are common devices in modern integrated circuits. MOS is the acronym for "metal-oxide-
10 semiconductor," which now encompasses both metal and polysilicon as the gate material.

The basic components of an MOS transistor are: a source that serves as a current input; a drain that serves as a current output; a channel that selectively couples the
15 source and the drain; and a gate that controls the conductivity of the channel. When a forward voltage bias is applied between the source and the drain, the current flowing from the source to the drain, through the channel, is controlled by the gate voltage.

20 A halo implant, also called a "pocket implant," has been used to reduce "punch through", i.e., to limit lateral diffusion of the source and the drain dopants in MOS transistors. The halo implant is of the conductivity type opposite to that of the source and drain of the MOS device.
25 Like the source/drain implant, it is performed after the gate is defined and before the source/drain diffusion. Due to the masking effect of the gate, the halo implant peak

concentration is near the source/drain region. Away from the source/drain edge, under the gate, the depth of the peak halo concentration falls quickly.

5   To provide a halo implant which extends slightly under the gate, high angle ion implantation is used to form the halo implant. Typically, the ion flux is at an implant angle of approximately 30°, or greater, from normal (perpendicular) to the substrate.

10   With close packing of devices, shadowing tends to occur due to the large angle employed. Shadowing results from the interference of an adjacent device gate, or other raised feature, with the path of the large angle, ion flux. To avoid the implant shadowing, devices may be spaced further apart. This results in fewer devices being
15 accommodated on a wafer.

               There remains a need for a method of halo implantation which overcomes the above-referenced problems, and others.

SUMMARY OF THE INVENTION

20   In accordance with aspects of the present invention, an MOS device and its method of fabrication are provided. The method includes forming a gate structure with an upper layer of a hard mask material on a substrate. The method further includes etching the hard mask material to
25 remove a portion of the hard mask material and form a contoured mask on the gate structure. The contoured mask varies in thickness across the gate structure. Further, the method includes implanting a halo dopant through the contoured mask into the substrate to form a halo implant.

30   In accordance with another aspect of the present invention, a method of forming a MOS transistor is provided. The method includes growing a gate oxide layer on a surface of a silicon substrate and depositing a layer of polysilicon on the gate oxide layer. The method further includes
35 forming a layer of a hard mask material on an upper surface of the polysilicon and anisotropically etching the hard mask

material to contour the hard mask material. A halo dopant is then implanted into the substrate through the contoured hard mask material at an implant angle which is generally perpendicular to the surface of the substrate to form a halo
5   implant.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention is best understood from the following detailed description when read with the accompanying drawing FIGURES. It is emphasized that various
10   features are not necessarily drawn to scale.

FIGURE 1 is a schematic elevational view of a portion of an MOS transistor in accordance with the present invention;

FIGURE 2 is an elevational view of a wafer during
15   preparation of the transistor of FIGURE 1, prior to etching of gate oxide and polysilicon layers;

FIGURE 3 is an elevational view of a wafer during preparation of the transistor of FIGURE 1, following etching of gate oxide and polysilicon layers and prior to etching of
20   a hard mask;

FIGURE 4 is an elevational view of a wafer during preparation of the transistor of FIGURE 1, following etching of a hard mask; and

FIGURE 5 is an elevational view of a wafer during
25   preparation of the transistor of FIGURE 1, following implant of a halo dopant.

DETAILED DESCRIPTION OF THE INVENTION

For purposes of illustration, the invention will be described in the context of MOS devices. However, it
30   should be readily appreciated that the invention is applicable to semiconductor devices in general, such as MOSFETS, HEMTs, BiCMOS, thin film transistors, and the like.

An MOS transistor 10 is illustrated in FIGURE 1. The transistor shown is a p-channel transistor, although it
35   should be readily appreciated that n-channel transistors may

be similarly formed, with the conductivities of the dopants reversed.

The p-channel transistor **10** is formed, for example, on and within a semiconductor substrate **12** which forms part of a wafer **W** on which many of such devices may be formed. Suitable substrates are silicon wafers, prepared, for example, by the Czochralski technique. The major surface of the wafer is not critical and any suitable orientation, such as a ⟨100⟩, ⟨110⟩, or ⟨111⟩, is useful. Other substrate materials, such as gallium arsenide, may also be employed.

Defined within substrate is a lightly doped n-well region **14**, within and over which the n-channel transistor **10** is formed. The n-well extends downward from the semiconductor substrate surface into its bulk. Alternatively, the bulk semiconductor may be doped, obviating the need for a separate well region. A dielectric film **16** covers the transistor. Contact and via structures **20** provide for electrical connections between transistor **10** and other transistors and devices of an integrated circuit (not shown).

The MOS device **10** comprises a gate **22**, and source **24**, and drain regions **26**. The source serves as a current input and the drain as a current output. A gate oxide **28**, such as a layer of grown silicon oxide, is disposed on the surface **29** of the wafer, serves to facilitate the projection of electric fields into a channel **30**, defined beneath the gate oxide which selectively couples the source and drain regions **24,26**. Channel **30** is controlled by the voltage applied to gate **22**. Thus, when a forward voltage bias is applied between the source and the drain, the current flowing from the source to the drain, through the channel, is controlled by the gate voltage.

The gate **22** may comprise a doped polysilicon layer **32**, disposed on the gate oxide, with an upper layer of tungsten silicide **33** thereon. Alternatively, the gate may

- 5 -

comprise a polysilicon layer without an upper tungsten silicide layer, or a layer comprising polycide.

In the case of a p-channel transistor, the source is p-type, the channel is p-type, and the drain, p-type. When an electrical bias is applied across the transistor, an inversion occurs in the channel. Source 24 preferably includes a lower, heavily doped source section 34 and an upper, lightly doped source section 36, adjacent the channel. Similarly, drain 26 preferably includes a heavily doped, lower drain section 38 and a lightly doped, upper drain section 40. The p-channel transistor 10 also includes a first, lightly-doped n-type halo implant region, or pocket implant 42 adjacent the source (hereinafter the source halo implant), and a second, lightly-doped n-type halo implant region 44 adjacent the drain 26 (hereinafter the drain halo implant).

It should be readily appreciated that n-channel transistors may be similarly formed, but with an n-type source 24, an n-type channel 24, an n-type drain 26, and lightly-doped p-type halo implants 42,44 (i.e., of a conductivity which is the opposite of the source and drain). CMOS devices may have both n-channel and p-channel in adjacent regions of a substrate.

The source and drain halo implant regions 42, 44 reduce diffusion from the source 34 to the drain 38 and thus contribute to an enhanced breakdown voltage for the transistor 10. Each halo implant 42, 44 has a side-wall section 46, 48, respectively, which extends vertically downward adjacent the channel 30, and either just outside or slightly under the gate 22. This portion of the halo implant region blocks the movement of ions from the source into the channel or from the drain into the channel. A lower section 50,52 of each of the halo implants may extend under the heavily doped source and drain regions, respectively, from the lower end of the vertical sidewall section 46, 48. Under the gate 22, the halo implant falls

away rapidly, so that the length of the channel is substantially unaffected by the halo implants.

A typical integrated circuit employing the MOS device of the present invention has a substrate 12 which is nominally 650 μm thick. The polysilicon gate is about 0.16 μm in length and 0.2-0.4 μm thick. The gate oxide is about 26 Å thick, or less. It will be readily appreciated, however, that the invention is also applicable to larger or smaller devices.

For 0.16 μm technology, the average dopant concentration of the well region 14, excluding the source, drain and halo implants, is between about 1×10^{14} and 5×10^{16} atoms/cm^3. The dopant concentrations for the source region 34 and the drain region 38 are preferably between about 1 and 2×10^{20} atoms/cm^3 for the heavily doped regions and between about 10^{19} and 8×10^{19} atoms/cm^3 for the lightly doped regions. The channel region 30 has a dopant concentration of less than about 1×10^{16} atoms/cm^3, more preferably, between about 1×10^{14} and 1×10^{16} atoms/cm^3, most preferably, about 1×10^{15} atoms/cm^3. Dopant concentration in the halo implants 42, 44 is preferably at least 1×10^{16} atoms/cm^3, more preferably, between about 1×10^{17} and 1×10^{18} atoms/cm^3. As the dopant concentrations generally vary inversely with device size; as device size increases, the dopant concentration increases. Thus, the recited concentrations may increase for larger device sizes.

Formation of the gate oxide 28, polysilicon layer or layers, 32, 33 and lightly doped and heavily doped regions 34, 36, 38, 40 may be achieved by conventional integrated circuit fabrication methods. For example, the following process may be used. First the well 14 is formed. An n-type implant defines the n-well 14 with somewhat higher concentration than substrate 12.

Field oxide 56 (silicon dioxide) is then grown to a thickness of about 0.3 μm on the silicon wafer surface. Alternatively, shallow trench isolation is used to isolate the devices.

A voltage adjust implant is performed while the n-well tub is defined. Dopant atoms, such as boron, are implanted into an upper region of the n-well **14** in the region at which the gate is to be formed to define a p-channel implant as shown in FIGURE 2. This implant is referred to as the threshold adjust implant and is used to set the threshold voltage at the gate. Preferably, the boron threshold adjust implant provides a nominal turn on (about 0.4 volts for 0.16 μm technology) for to-be-defined channel **24** in both PMOS and NMOS devices.

The gate **22** layers are then deposited. Specifically, an oxide for the gate oxide layer **28** is grown over the substrate surface, for example, by thermal oxidation. Polysilicon for the gate **22** is then deposited over the gate oxide. Various techniques, such as physical deposition, chemical vapor deposition, or epitaxial growth, may be used to perform this step. Preferably, an upper layer **34** of tungsten silicide, is then deposited over the polysilicon, by, for example, sputtering.

A layer **60** of a hard mask material is then deposited on an upper surface of the gate -i.e., on an upper surface **62** of the tungsten silicide layer **34**. The mask material is one which is etched selectively over polysilicon or silicon. The hard mask material is preferably an oxide, such as silicon oxide, or silicon nitride. However, other oxides which are deposited by chemical vapor deposition (CVD) may also be used. The silicon oxide may be formed, for example, by LPCVD or PECVD using a plasma comprising tetraethylorthosilicate (TEOS), and optionally hydrogen, to a thickness of between about 500 and 2000Å, preferably, about 1000-1500Å.

The gate is then defined by patterning the hard mask. For example, a layer of photoresist **64** is deposited on the area of the gate, as shown in FIGURE 2 and then the unwanted hard mask material is removed by reactive ion etching (RIE) using a halogenous gas as a source of ions, such as CHF_3, CF_4, a mixture thereof, or the like. The

remaining photoresist material is then stripped to expose an upper surface **66** of the hard mask. The unwanted portions of the tungsten silicide layer **33**, the polysilicon layer **32** and the gate oxide are then etched away to define the gate,

5   using the hard mask material as a mask. RF plasma etching with an HBr-based plasma is suitable for this step.

Doping of the lightly doped source and drain regions **34, 38** can then take place. For example, boron may be implanted at moderate concentration and low energy

10 (around 15 KeV) to create a p-type lightly doped drain implant in n-well **14**. Alternatively, the order of the lightly doped drain implant and the halo implant may be reversed.

The hard mask layer **60** is then etched with a

15 blanket anisotropic etch, as indicated by arrows **E** in FIGURE 3. The anisotropic etch may be carried out in a reactive ion etch (RIE) reactor, using argon. A preferred etch is carried out at about 1 mTorr.

Because of the anisotropic nature of the etch, the

20 hard mask **60** is not etched equally across its width. The anisotropic etch leaves the hard mask with a contoured, cap shape, as shown in FIGURE 4. The curved profile of the cap **72**, thus formed, has a maximum thickness (height **h** above the surface **62** of the polysilicon layers) at a mid-point **74** of

25 the mask material and a minimum thickness at or adjacent to the side edges **76** of the gate.

The contour of the contoured mask influences the depth profile of the subsequent halo implant under the gate structure. The thickness of the mask material may be

30 selected such that the maximum height **h** of the mask cap **72**, following etching, is approximately equal to, or slightly greater than, the depth of the lightly doped regions adjacent the gate. For example, if the lightly doped regions are to be about 1000Å in depth adjacent the gate,

35 the maximum height **h** of the mask cap **72** is also preferably about 1000Å, after etching.

The amount of mask material **70** removed in the blanket anisotropic etch can also be used to determine how far under the gate the halo implant **42, 44** extends. For example, the mask cap **72** of FIGURE 4, does not extend fully
5 to the periphery **76** of the gate. A width **w** is free of mask material **70**. This type of mask can be used to form a halo implant region **42, 44** that extends under the gate by an amount roughly equal to the width **w**, and then rapidly trails off, following the profile of the cap **72**. A mask **70** which
10 extends fully to the periphery of the gate can be used to create a shallower under-the-gate halo implant.

The shape of the stack (polysilicon and mask cap **22,72**) is then used to create a correspondingly-shaped halo implant **42, 44** in the substrate using an halo implant at an
15 implant angle which is normal (perpendicular), or close to normal, to the surface of the substrate. The implant is preferably at medium energy (nominally around 90 KeV phosphorus in an n-channel transistor, or around 65 KeV for boron in a p-channel transistor).

20 The depth of penetration of ions of a given energy is dependent on the thickness of the material through which the ions must travel. The present invention takes advantage of the contour of the cap **72** to provide a varying thickness through which the ions must travel, to implant the halo
25 dopant at a similarly contoured depth. The energy of the implant material, phosphorus, for example, in the case of the n-well, is selected, so that the ions have sufficient energy to penetrate through the stack to the substrate beneath only in a region **46, 48** adjacent the edges **76** of the
30 stack, where the hard mask cap **72** is thinnest or absent. In the center **74** of the stack, where the hard mask **72** is thicker, the ions do not have sufficient energy to penetrate to the channel region **30** below.

The profile of the vertical portion **46,48** of the
35 halo implant, adjacent the channel **30**, is thus generally defined by the profile of the mask material, as shown in FIGURE 5.

As will readily be appreciated, p-well transistors may be similarly halo implanted and lightly doped source and drain regions formed.   For example, phosphorus may be implanted at moderate concentration and low energy, around

5   25 KeV, to create an n-type lightly doped drain implant, then boron is implanted at high energy, nominally 65 KeV into the p-well to form a p-halo implant.   CMOS devices may be implanted in both the n and p-wells by first masking, for example, the n-well, implanting the p-well, then removing

10 the mask from the n-well and masking the p-well, and finally, implanting in the n-well.

As will be readily understood other steps in the processing may be included as known in the art.   For example, a pad oxide may be grown on the structure to

15 protect the substrate surface prior to the halo implant.

After the halo implant is created, dopant atoms are implanted into regions 36 and 38 to form the heavily doped regions of the source and drain.   The wafer W, which may have many such devices formed thereon is then annealed

20 to repair the implant damage (restore the silicon lattice structure) and to activate the dopants (placing the dopant atoms on vacant sites).   Finally, metal 20 and one or more intermediate layers 16 of an insulation material, such as a dielectric material of doped silicon dioxide are formed.

25 The metal provides the necessary connections with other devices on the wafer while the doped oxide serves as an intermediate dielectric to isolate the metal interconnect level from the polysilicon 32, 33.   The steps for fabricating an actual MOS transistor involves many more

30 steps.   These steps are well known and are described in S.M. Sze, "VLSI Technology," 2nd. ed., New York: McGraw-Hill (1988).

While it is preferable to use a perpendicular halo implant (i.e., an implant angle at about 0°, and preferably

35 no more than 7° relative to an axis x perpendicular with the substrate surface 29), for forming symmetrical halo implants adjacent both the source 34 and drain 38 simultaneously, the

implant angle could obviously be varied if desired to selectively halo implant either the source or the drain. For, example, the implant angle may be 7°-10° or more from perpendicular, without resulting in shadowing by adjacent

5   devices, depending on the pitch and dimensions of the devices. Where device spacing is not critical, the angle of implant may be as much as 30°, although at such large angles, the halo implant will extend further under the gate than may be desirable for some applications.

10   The method of halo implantation thus described allows implantation at shallow, i.e., generally perpendicular implant angles, which permits closer spacing of transistors on a wafer, without appreciably increasing processing time and thermal requirements.

15   The invention has been described with reference to the preferred embodiment. Obviously, modifications and alterations will occur to others upon reading and understanding the preceding detailed description. It is intended that the invention be construed as including all

20   such modifications and alterations insofar as they come within the scope of the appended claims or the equivalents thereof.